WEEKLY REPORT

Supaporn Lonapalawong(卢金璇)

2019/07/08-2019/07/14

本周工作

- 电网项目
  - 阅读 markov chain
  - 研究 cascading failure 系统
- 南网项目
  - 暂时没有工作
- 把电网运行状态可视化综述做 ppt 准备 script
- 阅读 4 篇论文
- 了解一下《R You Ready to Python? An Introduction to Working with Land Remote Sensing Data in R and Python》
- 因周五周六房间需要维修，没来实验室。
- 周三-周四早上去听数据可视化基础的课
-

下周工作计划

- 继续阅读论文

Abstract

[1]Remotely sensed measures of productivity are frequently used to characterize global agriculture and vegetated ecosystems, and are often downscaled to describe local, remote areas where finer spatial and temporal resolution data are regularly unavailable. While data errors may propagate throughout any analytical procedure, those that are missed during delivery and preliminary data mining require more attention. Here, a collection of formerly and presently available global remote sensing products are compiled to demonstrate the temporal and geographic breadth of remote sensing uncertainty. Vegetation productivity measures are invaluable for monitoring global health, but erroneous estimates that go unrecognized may result in serious policy mistakes. It is eminently clear that generalizable and accessible apriori methods for anomaly detection are lacking and urgently needed so that data errors are recognized before public delivery and before widespread use. Simple yet effective statistics such as themodified Z-score, Tukey's outliers, and Geary's Care leveraged here to identify, locate, and visualize the types of outliers that remote sensing data users may elect to omit or correct. Contributing to the growing ensemble of Google Earth Engine methodologies, we propose this generalizable method of detecting spatial outliers for remote sensing error management by users across scientific domains.

[2]Today's artificial intelligence still faces two major challenges. One is that, in most industries, data exists in the form of isolated islands. The other is the strengthening of data privacy and security. We propose a possible solution to these challenges: secure federated learning. Beyond the federated-learning framework first proposed by Google in 2016, we introduce a comprehensive secure federated-learning framework, which includes horizontal federated learning, vertical federated learning, and federated transfer learning. We provide definitions, architectures, and applications for the federated-learning framework, and provide a comprehensive survey of existing works on this subject. In addition, we propose building data networks among organizations based on federated mechanisms as an effective solution to allowing knowledge to be shared without compromising user privacy

[3]Subsequence clustering of multivariate time series is a useful tool for discovering repeated patterns in temporal data. Once these pat- terns have been discovered, seemingly complicated datasets can be interpreted as a temporal sequence of only a small number of states, or clusters. For example, raw sensor data from a fitness-tracking application can be expressed as a timeline of a select few actions (i.e., walking, sitting, running). However, discovering these patterns is challenging because it requires simultaneous segmentation and clustering of the time series. Furthermore, interpreting the resulting clusters is difficult, especially when the data is high-dimensional. Here we propose a new method of model-based clustering, which we call Toeplitz Inverse Covariance-based Clustering (TICC). Each cluster in the TICC method is defined by a correlation network, or Markov random field (MRF), characterizing the interdependencies between different observations in a typical subsequence of that cluster. Based on this graphical representation, TICC simultaneously segments and clusters the time series data. We solve the TICC problem through alternating minimization, using a variation of the expectation maximization (EM) algorithm. We derive closed-form solutions to efficiently solve the two resulting subproblems in a scalable way, through dynamic programming and the

alternating direction method of multipliers (ADMM), respectively. We validate our approach by comparing TICC to several state-of-the-art base- lines in a series of synthetic experiments, and we then demonstrate on an automobile sensor dataset how TICC can be used to learn interpretable

[4]Time Series data has become ubiquitous thanks to affordable edge devices and sensors. Much of this data is valuable for decision making. In order to use these data for the forecasting task, the conventional centralized approach has shown deficiencies regarding large data communication and data privacy issues. Furthermore, Neural Network models cannot make use of the extra information from the time series, thus they usually fail to provide time series specific results. Both issues expose a challenge to large-scale Time Series Forecasting with Neural Network models. All these limitations lead to our research question:Can we realize decentralized time series forecasting with a Federated Learning mechanism that is comparable to the conventional centralized setup in forecasting performance?In this work, we propose a Federated Series Forecasting framework, resolving the challenge by allowing users to keep the data locally, and learns a shared model by aggregating locally computed updates. Besides, we design a hybrid model to enable Neural Network models utilizing the extra information from the time series to achieve a time series specific learning. In particular, the proposed hybrid outperforms state-of-art baseline data-central models with NN5 and Ericsson KPI data. Meanwhile, the federated settings of purposed model yields comparable results to data-central settings on both NN5 and Ericsson KPI data. These results together answer the research question of this thesis.

[5]The LP DAAC operates as a partnership between the U.S. Geological Survey (USGS) and the National Aeronautics and Space Administration (NASA) and is a component of NASA's Earth Observing System Data and Information System (EOSDIS). Data specialists, system engineers, user service representatives, and science communicators work in collaboration to support LP DAAC activities.The LP DAAC processes, archives, and distributes land data products to hundreds of thousands of users in the earth science community. Our land data products are made universally accessible and support the ongoing monitoring of Earth's land dynamics and environmental systems to facilitate interdisciplinary research, education, and decision-making.

Reference

[1] Peter BG, Messina JP. Errors in Time-Series Remote Sensing and an Open Access Application for Detecting and Visualizing Spatial Data Outliers Using Google Earth Engine. *IEEE J Sel Top Appl Earth Obs Remote Sens*. 2019;12(4):1165-1174. doi:10.1109/JSTARS.2019.2901404

[2] Yang Q, Liu Y, Chen T, Tong Y. Federated Machine Learning. *ACM Trans Intell Syst Technol*. 2019;10(2):1-19. doi:10.1145/3298981

[3] Hallac D, Vare S, Boyd S, Leskovec J. Toeplitz inverse covariance-based clustering of multivariate time series data. *IJCAI Int Jt Conf Artif Intell*. 2018;2018-July:5254-5258.

[4] Neto PSGDM, Petry GG, J ARL, Ferreira TAE. for Time Series Forecasting. 2009:2230-2237.

[5] https://lpdaac.usgs.gov/resources/e-learning/